

With the rise in popularity of the modern lifestyle globally, settling in developed nation has been a common nature for all sectors of people. So, analyzing the population of the most common cities and choosing the best possible scenario seems to be a right topic for the project.

The Data is collected from the Wikipedia about the cities of the three states. Now, the data collected is queried for the location and on the popularity of the venue for the proper classification of the cities to be chosen by the personals.

Srinivasan Sugumar

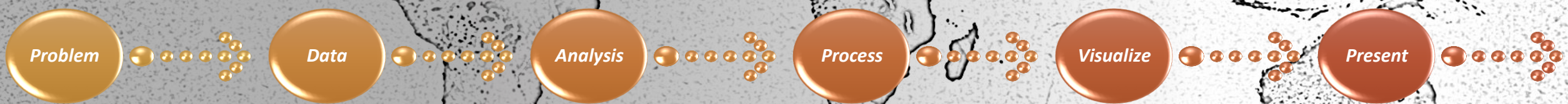
Illinois

New York

California



The ALTERNATIVE CITIES TO LIVE IN THE US FROM THE TOP 3 STATES OF POPULATION; IBM Data Science Capstone Project consists of the following processes.



Illinois

New York

California

Over the past decade, New York, California and Illinois are the fastest growing states in the United States. These are the most populous incorporated states of the United States including designations, city, town, village, borough and municipality.

Problem Data

	2019rank	City	State[c]	2019estimate	2010Census	Change	2016 land area	2016 land area.1	2016 population density	2016 population density.1
0	1	New York[d]	New York	8336817	8175133	+1.98%	301.5 sq mi	780.9 km2	28,317/sq mi	10,933/km2
1	2	Los Angeles	California	3979576	3792621	+4.93%	468.7 sq mi	1,213.9 km2	8,484/sq mi	3,276/km2
2	3	Chicago	Illinois	2693976	2695598	-0.06%	227.3 sq mi	588.7 km2	11,900/sq mi	4,600/km2
3	4	Houston[3]	Texas	2320268	2100263	+10.48%	637.5 sq mi	1,651.1 km2	3,613/sq mi	1,395/km2
4	5	Phoenix	Arizona	1680992	1445632	+16.28%	517.6 sq mi	1,340.6 km2	3,120/sq mi	1,200/km2

We are collecting the data sources of the towns, cities and venue details about the three states and analyzing them to find and guide a better place among these three states that have all the satisfying amenities like better workplace, living space and settling with family

Problem



Illinois

New York

California

Sources of Data



To obtain a list of cities of these states, we'll scrape Wikipedia for a list of cities in New York, California and Illinois based on the 2019 estimated population

Location Data



Mapquest Geocoding API used to tag the location of the towns and cities of the three states (latitude/longitude)

Venue Details and Data



Foursquare API venue recommendation used to obtain a list of the most popular venues for each towns and cities based on the Mapquest API's latitude and longitude tags



Illinois

New York

California

Analysis



We will scrape Wikipedia for a list of cities in New York, California and Illinois. The lists of cities on those pages are structured in tables, so we can easily use Pandas to read in the HTML table and convert it to a dataframe.

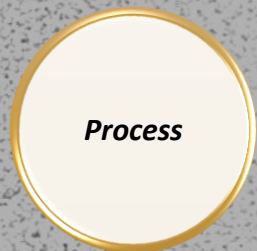


We are using Python and Packages such as pandas, sklearn and folium to find the solution for the stated problem. We will cleanse the data for cities those does not have any venue data.

The k-means clustering algorithm will be used for multiple iterations in order to find the best fitment of data for the classification.



Finally folium will be used to plot the data into a map to understand how the classification of clustering are shown visually. This will be the baseline data for our final presentation or conclusion and recommendation.



Illinois

New York

California

Selection of Venues

01

Based on the current population we decided to filter more than 10 venues from each towns and cities and top 5 venue categories for each towns and cities

Methodology (Algorithm)

02

We are using K-Means clustering unsupervised algorithm for the classification of data (Venues)

Cluster

03

The K-Means cluster centroids 7 are the most appropriate for us to analyze the data

Illinois

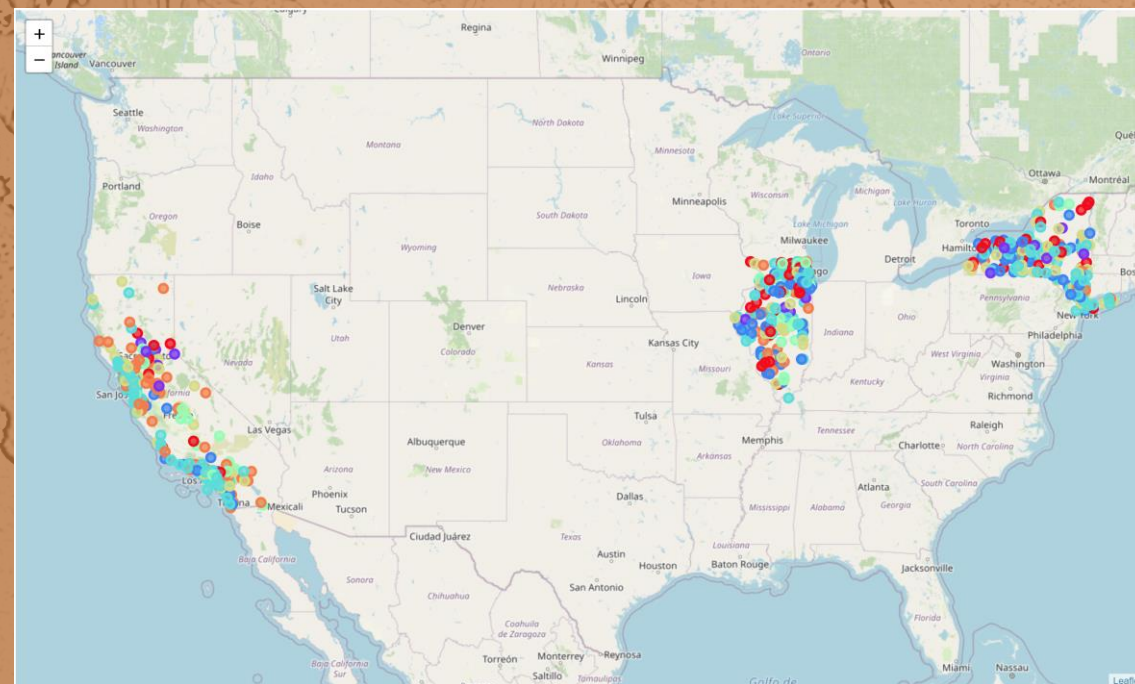
New York

California

Visualize

Classification of Visualization

Outliers
Vacation destinations
Restaurant towns and cities
Coffee shops towns and cities
Fast food towns and cities
More Restaurants





Illinois

New York

California

Present

Presentation - Conclusion and Recommendations

Conclusion

Total towns and cities from the three states - 24,366

Towns and cities more than 10 venues - 18,127

Cities with more than 10 venues - 644



Unique Venue Categories - 462

Recommendation

It is clearly known that people, investors and real estate persons can focus on these towns and cities. We found that these cities are having all the needed facilities with the available data from the Foursquare API. We would like to have more data and analyze to fine tune the model for better results.

